PAPER

Review Rating Prediction on Location-Based Social Networks Using Text, Social Links, and Geolocations

Yuehua WANG^{†a)}, Student Member, Zhinong ZHONG^{†b)}, Anran YANG[†], and Ning JING[†], Nonmembers

SUMMARY Review rating prediction is an important problem in machine learning and data mining areas and has attracted much attention in recent years. Most existing methods for review rating prediction on Location-Based Social Networks only capture the semantics of texts, but ignore user information (social links, geolocations, etc.), which makes them less personalized and brings down the prediction accuracy. For example, a user's visit to a venue may be influenced by their friends' suggestions or the travel distance to the venue. To address this problem, we develop a review rating prediction framework named TSG by utilizing users' review Text, Social links and the Geolocation information with machine learning techniques. Experimental results demonstrate the effectiveness of the framework.

key words: review rating prediction, location-based social networks, multi-class classification, ensemble algorithm

1. Introduction

In recent years, online review sites and location-based services have been developed rapidly. With more and more local businesses joining in, online review sites such as Foursquare, Yelp, etc., have attracted more and more users. People can write reviews, rate businesses, connect with other users, and “check in” at places. They can also benefit from these websites by getting advice for traveling, reducing the time they take choosing where to eat and enjoy high quality services of high rated businesses. These online reviews function as the “online word-of-mouth” for consumers to choose between similar products [1]. Consumer purchase decisions, product sales and business revenues are also greatly impacted by the reviews. Review rating prediction is of much importance for sentiment analysis and business intelligence, because it enables the market to estimate how satisfied a customer will be with a product [2]. Horrigan [3] pointed out that consumers are willing to pay from 20% to 99% more to buy a product whose rating is 5-star rather than 4-star. Companies have more fine-grained requirements for keeping track of product quality [4]; for example, they would want to know which aspects (food quality, ambiance, or free Wi-Fi, etc.) of services they provide attract consumers, and which disappoint them. Therefore, a lot of attention has been paid to sentiment analysis and

opinion mining from both industry and research communities [5]. The online review sites are effecting our lives more and more.

A typical restaurant review usually contains the reviewer's simple information, an overall star rating, date, review text and some feedbacks of other users (e.g., votes of usefulness). Take Fig. 1, for example, reviews of the same restaurant from different users.

These two reviews discuss several aspects of the same restaurant, such as food quality, drink taste, and waiting time, etc., and overall ratings for the restaurant. Different users have different biases, and always have different opinions towards the same businesses, thus giving different ratings. Alternatively, they may give the same overall ratings but consider different aspects of the same businesses.

1.1 Motivation

The volume of reviews grows so rapidly that it is becoming increasingly difficult for users to wade through numerous reviews to find the needed information. What's worse, there are fake ones among the great deal of reviews, making it even harder for consumers to identify their authenticity. Although much work has been done to alleviate this problem including extracting information from reviews, summarizing users' opinions, categorizing reviews according to opinion polarities, and extracting comparative sentences from re-

Star rating: 5 Date: 22/2/2017 Location: Stockton, United States Text: Oysters so fresh so good! Not much of an oyster person but when I tried them here I was so glad I did. ... And the clam chowder is amazing with a lot of stuff in there including lobster - not your typical chowder! Votes: Useful: 10, Funny: 1, Cool: 1
Star rating: 2 Date: 29/1/2017 Location: Log Angeles, United States Text: Saw this restaurant listed in ... I was really disappointed and found it not worth the 45min wait to sit down for. I personally don't think it's all its hyped up to be, the drink just has too much going on. It has a very bold horseradish taste amongst many other spices, topped with cilantro and bits of green onion... just not for me. ... I couldn't get over the overwhelming horseradish flavor, so unfortunately, I can't stay that I loved this place though I wanted to! Votes: Useful: 1, Funny: 1, Cool: 1

Fig. 1 Examples of reviews.

Manuscript received June 1, 2017.

Manuscript revised December 20, 2017.

Manuscript publicized June 1, 2018.

[†]The authors are with College of Electronic Science and Engineering, National University of Defense Technology, Changsha, 410073, China.

a) E-mail: yuehua.wang@hotmail.com

b) E-mail: Nong_nudt@hotmail.com

DOI: 10.1587/transinf.2017EDP7180

views, there are still problems that need to be addressed:

1. **The inconsistency of review text and star rating:** Review content and aspect rating play important roles in predicting review ratings [2], [6], [7]. The star rating given by a user usually corresponds to the related review text, but in many cases there is an inconsistency [6]. For example, a user may give a high rating when she gives a review of bad service, or vice versa. The relationship of the review text and the star rating is not clear due to many aspects (e.g., users' bias, services provided by businesses, waiting time, or even the weather). Product ratings are important because they affect how well a product will be adopted by the market.
2. **The influence of social links:** When we want to go to a restaurant, which one we will pick? Will we get advice from our friends and what kind of food might the friend recommend to us? How do our social links can affect our eating or shopping? McPherson et al. [4], [8] proposed that reviewers have mutual influence when they are connected in a social network. Moghaddam et al. [9] suggested that users are frequently affected by other's opinion in the decision process. Researchers have also achieved a preferable result using social links of users in review quality prediction [10].
3. **The impact of geolocations of places and users' current locations:** When someone is traveling, they may go to the famous points-of-interest, eat food in famous restaurants, or drink in famous bars. Thus, it seems that they would give high star ratings to these famous businesses. Besides this, when one goes to eat or play, one wants to go to places with a clean environment, good ambiance and good-quality services with nice neighbors. Hu et al. [11] pointed out that when a user visits a business, there is a good chance that they will walk by its neighbors, and there exists a weak positive correlation between a business's ratings and its neighbors' ratings, regardless of the categories of businesses.

To address these problems, we develop a unified framework to predict the deserved scores of reviews according its review text, the reviewers' personal information, as well as their social links and geolocations.

1.2 Brief Literature Review

Many efforts have been invested in information extraction from review mining, and these can roughly be divided into: sentiment polarity detection (thumbs up or thumbs down, or positive/negative) and review rating prediction (e.g., 1–5 stars). There are also researchs on point-of-interest recommendation utilizing reviews to enhance traditional collaborative filtering-based recommender systems.

1.2.1 Sentiment Polarity Detection

Tan et al. [12] use user-user relationships to improve the per-

formance of sentiment analysis for Twitter. Chung et al. [13] generate Bag-of-Sentiment-Concepts features to perform concept-level sentiment analysis. Khan et al. [14] present an entity-level sentiment analysis method by combining a lexicon and learning-based method to perform sentiment analysis on Twitter data. Devaraj et al. [15] detect sentiment polarity of free-form texts by using lexicon ensemble and lexicon pooling. Khan et al. [16] build a sentiment dictionary SentiMI by extracting sentiment terms from SentiWordNet and develop a complete framework using feature selection and extracting mutual information from SentiWordNet to improve sentiment polarity detection.

1.2.2 Review Rating Prediction

Pang and Lee [17] pioneer the review rating prediction field by regarding review rating prediction as a classification/regression problem. They build a rating predictor using the machine learning method under a supervised metric labeling framework. Qu et al. [18] propose a bag-of-opinion method to tackle this problem for Amazon.com reviews. They treat a triple (consisting of a root word, a set of modifier words, and one or more negation words) as opinion, assign a numeric score to the opinion, and predict a review's rating by aggregating the scores of opinions present in the review and combining them with a domain-dependent unigram model. Ochi et al. [19] improve the accuracy of rating prediction using feature words extracted from customer reviews to reduce the dimension of the feature vector. Mukherjee et al. [20] train a linear regression model to learn the user's specific aspect preference extracted from the review content, and predicts the overall review rating. Gao et al. [21] develop user-specific features to capture the user leniency. Wu and Ester [22] leverage user information with a combination between collaborative filtering and aspect-based opinion mining. Linshi [7] proposes a modified, semantic-driven LDA model to personalize the ratings based on the different topics extracted across different user reviews.

Most of the works are oriented to the traditional review site by utilizing the semantics of review texts. Some of them utilize user specific features to capture personalized review rating prediction, [7], [21], [23]–[28]. However, few of them deal in particular with review rating prediction on Location-Based Social Networks (LBSNs).

1.2.3 Point-of-Interest Recommendation with Content Enhancement

Zhang and Chow [29] propose a recommendation approach with regard to geographical correlations, social correlations and categorical correlations among users and POIs to recommend new POIs to users. They predict the scores of user unvisited POIs to recommend top K POIs with the highest score for users. In their another paper [30], they also develop an opinion-based POI recommendation framework by detecting tips polarities and integrating it with social links

and geographical information into a unified recommendation framework. They detect tips polarities by extracting opinion phrases, clustering opinions by aspect distances and detecting the polarities with a supervised naive Bayesian classifier. They use this framework to predict the scores a user would give to the POIs that he/she never visited before, and then get the top K scored POIs as the recommendation list. Gao et al. [33] consider the content an important role in recommending new POIs to users, and build model to generate user-interest and POI property content, as well as user sentiment indications into a unified recommendation framework.

POI recommendations with content enhancement can improve the performance of recommender systems since they can get users' interests as well as POIs' properties, which would generate much closer candidate POIs satisfying the users' needs. The rating prediction in recommendation systems is used for users to get potential scores that users would give to the unvisited POIs in the future, while review rating prediction in this paper is to predict the deserved scores correlated to a review text which is given by the reviewer after he/she has visited the POI.

In this paper, we carefully consider the special characteristics of LBSNs, and propose a review rating prediction frame, trying to find the real rating correlated to a review, by exploring the review text, users' social links, users' and businesses' geolocation. The contributions of our work are as follows:

- We propose a social-enhance DBSCAN to infer users' regular locations;
- We give a description of the correlation between the social links and the star ratings;
- We analyze the correlation between the distances and star ratings;
- We propose an ensemble learning-based framework utilizing users' review text, social links and geolocations to predict the star ratings of the reviews.

This paper is organized as follows: Sect. 2 first gives a description of the preliminarily review rating prediction, then analyzes the correlations of star ratings ~ distances and star ratings ~ social links, and finally describes the framework we propose, including the users' regular locations inference, social link influence modeling, geolocation influence modeling and the combining method. Section 3 gives an overview of the dataset we use, the baseline models we will compare with, and the settings of our experiments. Section 4 discusses the results of our experiments. Section 5 concludes the work and suggests some potential areas to delve into in the future.

2. Methodology

In this section, we describe the proposed machine learning method for review rating prediction. First we give the description to the review rating prediction problem in our consideration. Then we give an overview of our TSG method

before delving into detail. Finally, we describe the use of our method for review rating prediction.

2.1 Problem Definition

Here we give a description to the review rating prediction problem on LBSNs.

Definition 1 (Regular Location): A regular location of a user is a place that the user often checks-in around, and can be regarded as the user's current location in some cases. A regular location is denoted as $rl_u = (latitude, longitude)$. Each user only has one regular location.

Definition 2 (Review): A **review** is a comment written by a user u to an item i , associated with a review text, star rating, user identification, item identification, review time, and other user-interactive properties. A review can then be denoted as $d_{u,i} = \{review_id, user_id, item_id, review_text, rating_star, time\}$. The *text* is comprised of a set of words $\{w_1, w_2, \dots, w_n\}$.

Especially in the case of LBSNs, the items are always venues or POIs which have the geolocation labels (i.e., latitudes and longitudes).

Definition 3 (Social Link): There are two kinds of social links: if two users are friends on LBSNs, we call this **real-social-link (rsl)**, and if two users are neighbors, we call this **geo-social-link (gsl)**. By "neighbor", we mean that the regular locations of users are within a certain distance (e.g., 5km).

In order to mine social links of users, we construct a graph G with users as vertices and social links (rsl and gsl) as edges, and then we get two adjacency matrices M_{rsl} and M_{gsl} , where $m_{ij} = 1$ if user u_i has social link with user u_j , otherwise $m_{ij} = 0$.

Definition 4 (Review Rating Prediction (RRP)): Given a review $d_{u,i}$ written by user u to item i , user information, social links and corresponding geolocation information as the input, RRP aims at inferring the numeric rating (1~5 stars) of $d_{u,i}$.

2.2 Regular Location Inference

Regular location inference infers users' regular locations from their check-in records. On LBSNs, users usually check-in at places to show their attendance. Researches [30]–[33] have suggested that users' check-ins always have a spatial aggregation effect, that is a user may check-in many times at a few places. Intuitively, a user may check-in at two places most often: the home and the workplace. Thus, these two places are the most probable places that a user may appear. Figure 2 shows a typical user's check-in distribution. Therefore, we can infer a user's regular location considering the user check-in history. On the other hand, a user may have many interactions with their

Fig. 2 A typical user's check-in distribution. This shows that a user may check-in at some certain places many times, and at other places a few times.

Fig. 3 Model to infer users' regular location with respect to their check-in history and social links.

friends and may go to some places together. Thus, we can infer the user's regular location by their friends' check-ins. Our user regular location inference model is as shown in Fig. 3.

A user u 's check-ins are a set of places $\mathbf{B}_u$, and check-in times are $\mathbf{w}_u$. The user's friends' check-ins are $\mathbf{B}_f$, and check-in times are $\mathbf{w}_f$. Thus, we can get a check-in set $\hat{\mathbf{B}}$, and the check-in times $\hat{\mathbf{w}}$. Then we can perform weighted DBSCAN to cluster these places into several groups. According to [36]–[38], the check-in distances of two successive POIs are always within a certain range, typically 10km, which is approximately 0.1° of latitude/longitude. The diameter of the range of human activities on weekdays is about 20km. Thus we set $\epsilon = 0.1$ to perform DBSCAN on the latitude/longitude points. The minimum points in a cluster is set as 5, i.e., $\minPts = 5$. Then we pick the largest cluster $\mathbf{C}$ and weight $\mathbf{w}_c$ as basic places and check-in times, and can get the user's regular location with:

$$\hat{h} = \frac{1}{M} \sum_{i=1}^M w_c * C_i \quad (1)$$

where M is the number of points in the cluster $\mathbf{C}$, and C_i is the i^{th} point (longitude, latitude).

Table 1 Correlation between star ratings and distances. 'Estimated' is the estimated values of parameters β_i and $Pr(> |t|)$ is the p-values.

Distance (km)	Estimated	t value	Pr(> t)	Size
(0, 8000]	β_0	3.7784	2714.7	<2e-16
	β_1	-0.6338	-0.5630	0.5737
	β_2	0.4661	0.4140	0.6790
	β_3	-3.6461	-3.2370	0.0012
(0, 10]	β_0	3.7799	2311.4	<2e-16
	β_1	6.8470	6.1130	9.78e-10
	β_2	-6.6140	-5.9050	3.53e-09
	β_3	3.2174	2.8720	0.0041
(10, 100]	β_0	3.7768	1315.8	<2e-16
	β_1	-0.3125	-0.2710	0.7860
	β_2	-0.5995	-0.5200	0.6030
	β_3	0.6863	0.5960	0.5510
(100, 8000]	β_0	3.7613	547.7	<2e-16
	β_1	2.2698	2.1020	0.0355
	β_2	-0.4587	-0.4250	0.6709
	β_3	-1.1870	-1.1000	0.2715

2.3 Correlations

2.3.1 Star Ratings vs. Social Links

When traveling or visiting some businesses, users are always affected by their friends. They may tend to their friends for advices and recommendations. And after traveling or visiting the businesses, they will also be influenced by their friends when leaving the reviews to the places. For example, one would leave a negative comment to a restaurant even if he/she likes it but his/her friends do not. Intuitively, if a user has many friends, he/she would have a higher opportunity to be influenced by his/her friends. We first calculate the Pearson correlation coefficient between star ratings and social friends numbers.

2.3.2 Star Ratings vs. Distances

Star ratings are related to the geolocations of users and POIs. If a user is visiting a restaurant, he/she would like a neat environment rather than a dirty one, and he/she may prefer near ones to far ones. This is the underlying common senses of human activities and the start point of our research. We conduct polynomial regression on star ratings vs. distances between users and POIs of Yelp dataset (about 655143 pairs of star rating-distance records), finding that: **when users are traveling out-of-town, they would like to give higher star ratings than when they are in-town.**

We formulate the correlation as:

$$\hat{y} = \beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 d^3 \quad (2)$$

where $\hat{y}$ is predicted star ratings, d are distances between users and POIs, and $\beta_0, \beta_1, \beta_2, \beta_3$ are parameters to be evaluated.

We fit the model on 4 groups of data and the results are as shown in Table 1. Bold $Pr(> |t|)$ means it is statistically significant. We can see that the overall star ratings are related to the distance, but is weak. When the distance

Fig. 4 Review text feature generation flow.

are in 10km, the star ratings changes with the distance, and the trends in the range (10, 100] disappears. But when distances are larger than 100km, the star ratings will slightly increase along with the distance. This phenomenon is mainly because that users are familiar with the POIs that are near, and they would give rating as they please; while when traveling to other cities, they may visit the well-known places which usually have good reputations, and thus they would give higher star ratings.

2.4 Feature Construction

In this section, features of review text, social links and geolocations are constructed with different techniques respectively.

2.4.1 Review Text Features

The review text is an important information source for review rating prediction. It contains rich opinion information. However, the review text is in free form, and users may excessively use capital letters and punctuation marks to express strong opinions. The free-form text also contains many stop words, like “the”, “that”, “is” etc. Therefore, it is necessary to preprocess the review text in order to extract meaningful content. We use a unigram model and bigram model to extract features from the review text. We first create a dictionary of all the words and bigram phrases in the review corpus (We refer to a word or phrase as a term). Then a term-review matrix is constructed, where entry (i, j) is the frequency of occurrence of term i in the j 'th review. Then we remove the terms that occurred less than 1000 times. Finally, we apply the TF-IDF (Term Frequency - Inverse Document Frequency) weighting technique to this matrix to obtain the final text feature matrix $M_{n \times p_1}^{text}$. The review text feature generation flow is shown in Fig. 4. In the text feature matrix $M_{n \times p_1}^{text}$, each row represents a review text, and each column a term.

2.4.2 Social Link Features

There are two kinds of social links in our consideration (Fig. 5), as mentioned in Sect. 2.1. Users are likely to get advice from their friends for traveling and consuming. On many occasions, friends would do similar things: going out for dinner, going to bars or traveling to POIs, etc. As for

Fig. 5 User social links. User U and C, U and D are real friends and they have real social links (rsl), while U, A and B are neighbors, thus, they have geographical social links (gsl).

neighbors, they would also have good opportunities to go to such places together. In order to make use of a user's social link information, one simple approach is to use the average star ratings of users who have social links with the target user. Thus, we extend our representative review vector with more social features. We get the real-social-links (rsl) by the user edges in the dataset and geo-social-link (gsl) by selecting the top K nearest users to the target users. We use $K = 30$ in this paper. Finally, we build the social link feature by extracting profiles of the target users' friends (both rsl and gsl). Particularly, we get the number of friends, star ratings, average review text lengths, number of votes they get, the physical distances between target user and his/her friends as well as user attributes into the social link feature matrix $M_{n \times p_2}^{sl}$.

2.4.3 Geolocation Features

Researches [11], [28]–[33] have shown that users tend to visit nearby places more than farther-away ones. When a user is visiting somewhere, they may walk past its neighbor. This shows a slight human effect on users' rating behaviors. Users always consider the environment around them when consuming, and a good circumstance may lead to a high rating score, while bad one may lead to a low score. When traveling a long distance, users tend to give higher star ratings than for nearby locations. Figure 6 shows that, when the distances are long, there are more chances that a user gives a higher star rating (more positive ratings, e.g., ratings stars ≥ 3). Inspired by this observation, we designed the Geolocation-based model, taking user locations (i.e., the longitudes and latitudes of user regular locations), business locations (i.e., business longitudes and latitudes) and the distances between user and business as features and then form the geolocation feature matrix $M_{n \times p_3}^{geo}$. The user location is inferred with DBSCAN mentioned in Sect. 2.2.

2.5 Review Rating Prediction Framework

The review rating model is built with ensemble learning [34] algorithm. Figure 7 shows the framework.

First, a base learner library is built, in which there are L learners with different kinds of learning algorithms (e.g.,

Fig. 6 Star ratings vs. distances. (a) shows that most of the ratings occurred within a short distance. (b), (c) and (d) shows the users' star ratings under different distances.

Fig. 7 Framework of review rating prediction.

eXtreme Gradient Boosting (XGB), Gradient Boosting Machine (GBM) and Random Forest (RF), etc.). These learners take feature matrices (Review feature matrix, social feature matrix and geolocation feature matrix, respectively) as inputs. Cross-validation method is used when training base learners on original training set $X_{n \times p}$, and will generate a matrix $Z_{n \times L}$ of cross-validated predictions, where L is the number of base learners.

Then, the metalearner $\hat{\Phi}$ is constructed by one hidden layer neural network, and is trained on Z and the labels of original training set X .

Finally fit L models (one for each base learner) on original training set X , and the L models along with $\hat{\Phi}$ is the ensemble model, which is used to generate predictions on the test set.

In this paper, we use XGB, GBM, and RF as the learning algorithms, therefore, formulating nine base learners on three feature matrices. A hidden layer neural network is used since there is no need to use multi hidden layers on the cross-validated prediction matrix $Z_{n \times 9}$ with only nine features.

Table 2 Yelp overall dataset. #users, #reviews and #businesses are the number of users, reviews and businesses, respectively. #social links is the number of user friendships.

Dataset	#users	#reviews	#businesses	#social links
Yelp15	10880	655143	61291	150741

Table 3 Statistics of each group of Yelp review dataset.

Group	#users	#businesses	#social links	#reviews
{2010–, 2011}	3393	28232	34715	165363
{2011, 2012}	2764	27334	27445	114427
{2012, 2013}	3131	31597	36695	129457
{2013, 2014}	3699	34343	41819	147328
{2014, 2015}	4041	34749	40230	149677

3. Experiment

3.1 Experiment Settings

To evaluate the proposed review rating prediction framework, we conduct experiments on the Yelp review dataset[†]. We choose the users who have more than 20 reviews and more than 5 real-social-link friends. The description of the datasets is shown in Table 2.

We split the Yelp dataset into five groups according to the review years: {2010–, 2011}, {2011, 2012}, {2012, 2013}, {2013, 2014}, {2014, 2015}, where ‘2010–’ represents the year before 2010 (including 2004 to 2010) due to fewer reviews. For each group, the first part is treated as training set and the latter as a testing set. The detailed description of each group is given in Table 3. Figure 8 shows all histograms of the five groups star ratings. We can see

[†]https://www.yelp.com/dataset_challenge

Table 4 Experimental results for review rating prediction on five groups of a Yelp review dataset.

Method	{2010-, 2011}		{2011, 2012}		{2012, 2013}		{2013, 2014}		{2014, 2015}	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Majority	0.7760	1.1082	0.7760	1.1009	0.7820	1.0817	0.8280	1.1559	0.8790	1.2029
BOW-GBM	0.6950	1.0483	0.6750	1.0281	0.6680	0.9960	0.6200	0.9920	0.7170	1.1580
BOW-XGB	0.7020	1.0517	0.6380	0.9899	0.6850	1.0242	0.5840	0.9274	0.6470	1.0531
BOW-RF	0.7290	1.0728	0.7070	1.0644	0.6840	1.0208	0.7830	1.1362	0.7510	1.2317
TSG	0.6130	0.9471	0.5850	0.9225	0.6270	0.9524	0.5400	0.8809	0.6120	1.0198

Fig. 8 Distribution of star ratings of each group.

that most of the ratings are about 4 or 5 stars, and 4 stars are the most common.

We treat review rating prediction as an ordinal multi-class classification problem and conduct experiments on these groups and use MAE (Mean Average Error) and RMSE (Root Mean Squared Error) as the evaluation metrics to measure the divergences between predicted ratings and actual ratings.

$$MAE = \frac{1}{|\tau|} \sum_{i \in \tau} |r_i - \hat{r}_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{|\tau|} \sum_{i \in \tau} (r_i - \hat{r}_i)^2} \quad (4)$$

where $|\tau|$ is the size of the testing set and i represents the i^{th} review.

3.2 Baseline Methods

Comparative baseline methods for review rating prediction are described as follows:

- **Majority**: assigns the majority rating score in the training set to each review in the test dataset.
- **BOW**: bag-of-words (BOW) method, simply uses review text to form a TF-IDF matrix with respect to unigrams and bigrams. We build the predictor with Random Forest (RF), Gradient Boosting Machine (GBM) and eXtreme Gradient Boosting (XGB).

4. Results and Analysis

Table 4 shows the experimental results of the baseline methods and our method on the five group datasets. Our method, which uses review text, user social links and geolocations, is abbreviated as TSG.

Figure 9 shows the review rating prediction results for five groups of datasets. From these, we can see that the

Fig. 9 Review rating prediction results on each group of dataset. ‘maj’ is Majority-based method; ‘xgb’ is for BOW with XGB; ‘gbm’ is for BOW with GBM; ‘rf’ is for BOW with RF; ‘tsg’ is the proposed method in this paper.

performances of these methods are consistent on five group datasets and above all, our method TSG performs the best. Specifically, the Majority-based method performs poorest almost all the time as it does not capture any text-level or user-level information. However, the Majority-based method performs much better than random guess, in which each star rating have 20% opportunities. This is because the star ratings are imbalanced — they are all left-skewed, just as Fig. 8 shows.

Baselines based on BOW use the surface form of words in the review, and perform much better than the Majority-based method. However, they do not take into consideration social links and geolocations, thus missing lots of user-specific information. Among the three BOW-based algorithms, RF performs the least well, GBM better, and XGB the best. This is mainly because: RF randomly select features for all the decision trees, which are different from each other. RF finally average the trees together, which makes it much general, but less precise. GBM is an ensemble of weak decision trees, fits consecutive trees where each solves for the net loss of the prior trees, and the results of new trees are applied partially to the entire solution. This lead to a much precise prediction but less robust, since it’s easier to be over-fitting. XGB is based on the boosted trees and supports L1 and L2 regularization to control the model complexity, it uses second-order Taylor expansion of loss function for trees tuning. Above all, the proposed TSG method gives the best performance in all the five groups, since we

Fig. 10 Effect of Geolocations and Social links. ‘text only’ uses only review text; ‘text-geo-soc’ uses review text, geolocations and social links.

take into consideration the users’ geolocations and social links. Users’ social links provide user social influence, and prove the assumption that people will be influenced by their friends and neighbors. User geolocations provide users’ check-in history, and indicate users’ location biases. For example, some people prefer to travel, while some prefer to stay in a limited area, and they may give different star ratings to the same place.

We also conduct experiments with our TSG model to evaluate the effect of user geolocations and social links. Figure 10 shows the experimental results. After utilizing the review text, social links and geolocations, the model performs much better than just utilizing review text alone. Each group has a 5%+ improvement except for the first group.

5. Conclusions

In this paper, we introduce an ensemble method that incorporates user information, social links, and geolocation for review rating prediction. We treat review rating prediction as a multiclass classification problem and build a model to understand the review text content effect, social link effect and geolocation effect and combine these with an ensemble model. We conduct experiments on five groups of Yelp review datasets, and compare it against several baseline methods. Experimental results show that the proposed method has better performance than the baseline methods which only use textual semantics. However, there are some limitations: 1) The accuracy of the user regular location inference. User reviews are quite sparse in LBSNs, especially for Yelp. Users do not have many check-ins and their check-ins are quite sparse on both geolocations and time. Users do not check-in all the time, in fact, in most cases are they check-in just once. 2) Review texts are free-form in content, and many informal slang words and trending words are used frequently, which increases the difficulty to understand the meaning of a sentence. In the future, we will further mine the relations between review ratings and users’ geolocations,

social links and neighboring venues, and utilize other language models (e.g., LDA model) in order to better understand the semantics of the users’ review text.

References

- [1] C. Dellarocas, “The digitization of word of mouth: Promise and challenges of online feedback mechanisms,” *Management Science*, vol.49, no.10, pp.1407–1424, 2003.
- [2] Z. Jin, Q. Li, D.D. Zeng, Y. Zhan, R. Liu, L. Wang, and H. Ma, “Jointly modeling review content and aspect ratings for review rating prediction,” *Proc. 39th International ACM SIGIR Conf. on Research and Development in Information Retrieval*, Pisa, Italy, pp.893–896, 2016.
- [3] J.B. Horrigan, “Online shopping,” *PEW INTERNET & AMERICAN LIFE PROJECT*, 2008.
- [4] B. Wang, Y. Min, Y. Huang, X. Li, and F. Wu, “Review rating prediction based on the content and weighting strong social relation of reviewers,” *Proc. 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing*, San Francisco, California, USA, pp.23–30, 2013.
- [5] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol.2, no.1-2, pp.1–135, 2008.
- [6] M. Angioni, M.L. Clemente, and F. Tuveri, “Combining opinion mining with collaborative filtering,” *Proc. 11th International Conf. on Web Information Systems and Technologies*, pp.370–380, 2015.
- [7] J. Linshi, “Personalizing yelp star ratings: A semantic topic modeling approach,” *Yale University*, 2014.
- [8] M. McPherson, L. Smith-Lovin, and J.M. Cook, “Birds of a feather: Homophily in social networks,” *Annual Review of Sociology*, vol.27, no.1, pp.415–444, 2001.
- [9] S. Moghaddam and M. Ester, “ILDA: Interdependent LDA model for learning latent aspects and their ratings from online product reviews,” *SIGIR*, Beijing, China, pp.665–674, 2011.
- [10] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi, “Exploiting social context for review quality prediction,” *Proc. 19th International Conf. on World Wide Web*, pp.691–700, 2010.
- [11] L. Hu, A. Sun, and Y. Liu, “Your neighbors affect your ratings: On geographical neighborhood influence to rating prediction,” *Proc. 37th International ACM SIGIR Conf. on Research & Development in Information Retrieval*, Gold Coast, Queensland, Australia, pp.345–354, 2014.
- [12] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, “User-level sentiment analysis incorporating social networks,” *Proc. KDD*, pp.1397–1405, 2011.
- [13] J.K.-C. Chung, C.-E. Wu, and R.T.-H. Tsai, “Polarity detection of online reviews using sentiment concepts,” *Semantic Web Evaluation Challenge*, Communications in Computer and Information Science, vol.475, pp.53–58, Springer, 2014.
- [14] A.Z.H. Khan, M. Atique, and V.M. Thakare, “Combining lexicon-based and learning-based methods for twitter sentiment analysis,” *IJECSCSE*, pp.89–91, 2015.
- [15] M. Devaraj, R. Piriyani, and V.K. Singh, “Lexicon ensemble and lexicon pooling for sentiment polarity detection,” *IETE Technical Review*, vol.33, no.3, pp.332–340, 2016.
- [16] F.H. Khan, U. Qamar, and S. Bashir, “SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection,” *Applied Soft Computing*, vol.39, pp.140–153, 2016.
- [17] B. Pang and L. Lee, “Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales,” *Proc. 43rd Meeting of the Association for Computational Linguistics*, pp.115–124, 2005.
- [18] L. Qu, F. Informatics, G. Ifrim, and G. Weikum, “The bag-of-opinions method for review rating prediction from sparse text pat-

- terns," COLING, pp.913–921, 2010.
- [19] M. Ochi, M. Okabe, and R. Onai, "Rating prediction using feature words extracted from customer reviews," Proc. 34th International ACM SIGIR Conf. on Research and Development in Information Retrieval, Beijing, China, pp.1205–1206, 2011.
- [20] S. Mukherjee, G. Basu, and S. Joshi, "Incorporating author preference in sentiment rating prediction of reviews," Proc. 22nd International Conf. on World Wide Web, Rio de Janeiro, Brazil, pp.47–48, 2013.
- [21] W. Gao, N. Yoshinaga, N. Kaji, and M. Kitsuregawa, "Modeling user leniency and product popularity for sentiment classification," International Joint Conf. on Natural Language Processing, Nagoya, Japan, pp.1107–1111, 2013.
- [22] Y. Wu and M. Ester, "Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering," Proc. Eighth ACM International Conf. on Web Search and Data Mining, pp.199–208, 2015.
- [23] H. Xie and J.C.S. Lui, "Mathematical modeling and analysis of product rating with partial information," ACM Trans. Knowledge Discovery from Data, vol.9, no.4, pp.1–33, 2015.
- [24] D. Tang, B. Qin, T. Liu, and Y. Yang, "User modeling with neural network for review rating prediction," Proc. Twenty-Fourth International Joint Conf. on Artificial Intelligence, pp.1340–1346, 2015.
- [25] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," Conf. on Empirical Methods in Natural Language Processing, Lisbon, Portugal, pp.1422–1432, 2015.
- [26] G. Cosma and G. Acampora, "Neuro-fuzzy sentiment analysis for customer review rating prediction," Sentiment Analysis and Ontology Engineering, Studies in Computational Intelligence, vol.639, pp.379–397, Springer, 2016.
- [27] W. Zhang and J. Wang, "Integrating topic and latent factors for scalable personalized review-based rating prediction," IEEE Trans. Knowl. Data Eng., vol.28, no.11, pp.3013–3027, 2016.
- [28] G. Zhao, X. Qian, and C. Kang, "Service rating prediction by exploring social mobile users' geographic locations," IEEE Trans. Big Data, vol.3, no.1, pp.67–78, 2017.
- [29] J.-D. Zhang and C.-Y. Chow, "GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," SIGIR, pp.443–452, 2015.
- [30] J.-D. Zhang, C.-Y. Chow, and Y. Zheng, "OREC: An opinion-based point-of-interest recommendation framework," Proc. 24th ACM International Conf. on Information and Knowledge Management, New York, NY, USA, pp.1641–1650, 2015.
- [31] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," Proc. 34th International ACM SIGIR Conf. on Research and Development in Information Retrieval, pp.325–334, 2011.
- [32] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," CIKM, Shanghai, China, pp.739–748, 2014.
- [33] H. Gao, J. Tang, X. Hu, and H. Liu, "Content-aware point of interest recommendation on location-based social networks," Proc. National Conf. on Artificial Intelligence, vol.3, pp.1721–1727, 2015.
- [34] R. Polikar, "Ensemble based systems in decision making," IEEE Circuits Syst. Mag., vol.6, no.3, pp.21–45, 2006.
- [35] Z. Yuan, Y. Jiang, and G. Gid6falvi, "Geographical and temporal similarity measurement in location-based social networks," Proc. 2nd ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems, Orlando, Florida, USA, pp.30–34, 2013.
- [36] M. Ye, P. Yin, and W.-C. Lee, "Location recommendation for location-based social networks," Proc. 18th SIGSPATIAL International Conf. on Advances in Geographic Information Systems, San Jose, California, USA, pp.458–461, 2010.
- [37] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: A survey," Geoinformatica, vol.19,

no.3, pp.525–565, 2015.

- [38] G. Ference, M. Ye, and W.-C. Lee, "Location recommendation for out-of-town users in location-based social networks," Proc. 22nd ACM International Conf. on Information & Knowledge Management, San Francisco, California, USA, pp.721–726, 2013.

Yuehua Wang received his B.S. degree in Electronic Engineering from Tsinghua University, Beijing, China in 2010, and M.S. degree in Control Science and Engineering from Aviation University of Air Force, Changchun, China in 2012, and now is pursuing his Ph.D. in National University of Defense Technology (NUDT), Changsha, China. His research interests include recommender systems and online social network data analysis.

Zhinong Zhong received his B.S. degree in Applied Physics in NUDT in 1996, M.S. in Opto Electronics Engineering in 1999 and Ph.D. in Communication and Information Systems in 2003, respectively, in the same university. He is an associate professor of Department of Information Engineering, NUDT. His research interests include GIS and spatial data analysis.

Anran Yang majored in Geographical Information System (GIS), and received his B.S. and Ph.D. degree from Peking University, Beijing, China in 2010, and NUDT in 2016. His research interests include GIS and Volunteered Geographic Information (VGI).

Ning Jing received his B.S. degree in Communication and Information Systems in NUDT in 1983, M.S. in Signal and Information Processing in 1986 and Ph.D. in Computer Science in 1990, respectively, in the same university. He is a professor, director of Department of Information Engineering, NUDT. His research interests include GIS, database System, planning and decision support in spatial resources, spatial data analysis and visualization.